Clarification of some Matrix Theorems PhL 1.9.05

****

DISTINCT EIGENVALUES

Question: What do we know about an NxN matrix which has N distinct eigenvalues?

(1) The char poly exponents are all 1, char poly = min poly, and $a_i = b_i = 1$ in each of the N eigenmanifolds.
(2) Thus, the geometric multiplicity of the matrix is N, the same as the algebraic multiplicity.
(3) Thus, the set of N eigenvectors is linearly independent. This is what geomult means.
(4) Thus, if we create matrix X from these eigenvectors, we have $\det X \neq 0$ and X^{-1} exists.
(5) Thus, when we write $AX = X\Lambda$, we can rewrite that as $X^{-1}AX = \Lambda$. This says that we can diagonalize the matrix with a similarity.
(6) No claim is made that this similarity X is unitary. X would only be unitary if its columns were orthonormal, but all we know in general is that the columns are linearly independent. So we have:

Theorem: For a general NxN matrix A, eigenvectors of different eigenvalues are linearly independent, but are likely to not be orthogonal.

Theorem: If a general NxN matrix A has all different eigenvalues, it can be diagonalized by a similarity formed from the eigenvectors of A, but this in general will not be a unitary matrix because the eigenvectors, although linearly independent, are not necessarily orthogonal.

****

HERMITIAN WITH DISTINCT

Theorem: For a Hermitian NxN matrix A, eigenvectors of different eigenvalues are not just linearly independent, they are orthogonal. If we normalize them, then they are orthonormal.

Theorem: If the column vectors of a matrix are orthonormal, that matrix is unitary. (Just write $X^H X = 1$).

Theorem: If a Hermitian NxN matrix A has all different eigenvalues, it can be diagonalized by a *unitary* similarity formed from the eigenvectors of A.

****

HERMITIAN WITH REPEATED EIGENVALUES

Now let's consider a Hermitian matrix which does NOT have all different eigenvalues.

Theorem: Each eigenmanifold of a Hermitian matrix has geomult = algemult. It follows that the overall matrix then has geomult = algemult = N.

Theorem: The eigenvectors within an eigenmanifold of a Hermitian matrix can be made orthonormal by the usual G-S process.

Theorem: For a Hermitian NxN matrix, even if eigenvalues are *not* all different, one can *still* construct a set of N orthonormal eigenvectors as the diagonalizing matrix X, which is then unitary. Therefore, ANY Hermitian matrix can be diagonalized by a unitary similarity, regardless of the eigenmanifold structure.

****

SCHUR and BLOCK DIAGONAL THEOREMS

Theorem: (**Shur's Theorem**) Any NxN matrix can be brought to upper triangular form by a *unitary* similarity, and the N eigenvalues will then appear on the diagonal.

Theorem: (**Block Diagonal Theorem**) Any NxN matrix can be brought to block diagonal form by a similarity, where each block represents a distinct eigenmanifold of one eigenvalue. The dimension of each block is given by its algemult. This follows if nothing else from the direct sum theorem of Matthews. However, there is no claim made that this similarity is unitary! In general, it will NOT be unitary.

Very Simple Example with 1x1 blocks: Here is a (upper triangular and non-Hermitian) matrix B with two distinct eigenvalues. We want to get it into block diagonal form. Note that the eigenvectors are in fact linearly independent by theorem above, and we create a similarity matrix using these eigenvectors:

$B = \begin{pmatrix} \lambda_1 & a \\ 0 & \lambda_2 \end{pmatrix}$, eigenvectors are: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ for λ_1 and $\begin{pmatrix} \alpha_a \\ 1 \end{pmatrix}$ for λ_2 where $\alpha_a = a/(\lambda_2 - \lambda_1)$.

By doing a similarity with $S = \begin{pmatrix} 1 & \alpha_a \\ 0 & 1 \end{pmatrix}$ [inverse $\begin{pmatrix} 1 & -\alpha_a \\ 0 & 1 \end{pmatrix}$], we get $S^{-1} B S = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Note that the similarity here is NOT unitary because the columns are not orthogonal. This example shows that in general, the similarity which takes a matrix to block diagonal form will not be unitary.

Comments: If you start with an arbitrary matrix, you can get to block diagonal form as above with some non-unitary similarity. Once you are there, you can then take each block to upper triangular form by unitary similarity. Thus, the overall similarity which takes you from a general NxN matrix A to the block diagonal form of triangular blocks will not be unitary. Similarly, if you go on from there to the final Jordan block form, the overall similarity getting you to Jordan form is in general NOT unitary. As an example, see the Maple code for Matthew's case on his page 87 (corrected with -2 $\rightarrow$ -3). The similarity matrix P in this case has columns which are not orthogonal, and we do not have $P^T P = 1$.

****

NORMAL MATRICES

Theorem: Any square matrix A such that $[A,A^H] = 0$ can be diagonalized by a unitary similarity. Such a matrix is called a "normal matrix".

I have not proven this, but here are some examples:

(1) If A is Hermitian, then $A = A^H$ and then $[A,A^H] = 0$ trivially.

(2) If A is unitary, we know that $U^H = U^{-1}$ which tells us that $UU^H = 1 = U^HU$.

Thus, any Hermitian or any unitary matrix can be diagonalized by a unitary similarity.

Here is a web snippet on this subject:

> Normal Matrices
>
> A normal matrix is a square matrix which is commutative with its hermitian transpose. In fact, *a square matrix is normal if and only if it is unitarily diagonalizable*. Note that every unitary, hermitian or skew-hermitian matrix is a normal matrix. Moreover, let **A** be a normal matrix. We can determine whether it is unitary, hermitian or skew-hermitian by checking the eigenvalues:
>
> a) **A** is unitary $\Leftrightarrow$ $|\lambda| = 1$;
>
> b) **A** is hermitian $\Leftrightarrow$ $\mathrm{Im}(\lambda) = 0$;
>
> c) **A** is skew-hermitian $\Leftrightarrow$ $\mathrm{Re}(\lambda) = 0$;
>
> for every eigenvalue λ of **A**.

****

COMMENTS ON THINGS IN THE TOOMAS PDF NOTES.

1. Relation between Gauss Elimination and LU Decomposition. When you do Gauss elimination, you do row operations (ERO's) to make a matrix become upper triangular. If you can do this without any row swaps, then you end up with T = LA where A is your original matrix, and E is the composite of your EROs. As I show on page 3 of "matrix research", the non-swap EROs are all lower triangular with 1's on the diagonal, so you end up with this result:

(upper triangular) = (lower triangular with 1's on diagonal) * A or U = LA

Now think about this matrix L with 1's on the diagonal. We know that det L = 1, and L^{-1} exists. We know from our banded matrix theorem that L^{-1} is triangular same as L. And we know from our **Theorem M30** that the diagonal elements of L^{-1} are the inverses of those of L, so in our case they are still one. So the matrix L^{-1} has the same general form of L (unit diagonal, lower triangular). So rename L^{-1} as L', then we have shown that A = L' U and this is the LU decomposition. The crucial requirement is that no row-swaps are needed. If they are needed, we have to say A = PLU where P does the swaps.

2. Claim of Toomas. He says (page 95) that you can do the A = LU factorization if all "principle minors" don't vanish. This refers to all subdeterminants from size 1 to N-1, a pretty heavy restriction. He shows that if this is true, then you never end up with a 0 at a pivot point, and then you don't have to do row swaps.

Theorem: The product of a diagonal matrix D and an upper triangular matrix U will be an upper triangular matrix whose diagonal elements are the products of those of D and U. [Note that B = DU will have different off-diagonal elements from those of U.]

Proof: We know the product is triangular from our UT * UT = UT rule. Then just do an example and you see that the second fact is true.

Theorem: (The LDM Decomposition.) If all principle minors of A are non-vanishing, then we can write A = LDM where L is lower triangular with unit diagonal, and M is upper triangular with unit diagonal, and D is diagonal. All matrices of the decomposition are unique. { This is Proposition 6.1.1 page 142 Toomas }

Proof: From the assumption, we can say that A = LU where L is lower triangular with unit diagonal elements, and U is upper triangular with general diagonal elements. Let D by diagonal with the diagonal elements of U. Then define $M = D^{-1}U$. According to our previous theorem, this will clear out the diagonal elements so M will have a unit diagonal. Then we have A = LU = L(DM) = LDM, QED.

Theorem: (The LDM^T Decomposition.) If all principle minors of A are non-vanishing, then we can write A = M is upper triangular with unit diagonal, and D is diagonal.

Proof: This is just a restatement of the previous theorem.

Theorem. If A is *symmetric* and if all principle minors of A are non-vanishing, then the L = M in the LDM^T decomposition, so we then have $A = LDL^T$, where L is lower triangular with unit diagonal.

Proof: Write $A = LDM^T$ and then $A^T = MDL^T$. But since $A=A^T$, we must have L=M and $M^T = L^T$ because our theorem says that the decomposition is unique!

Definition: A matrix A is **positive definite** of the quadratic form $x^TAx > 0$ for all non-zero x.

Theorem: (1) All principle minors of a positive definite matrix are non-vanishing, and therefore you can always write the LDM^T decomposition for such a matrix. { See Toomas p 145 } (2) The diagonal elements of a positive definite matrix are positive. This follows at once since you can use x = unit vectors in the positive definite definition.

Theorem: Positive definiteness is invariant under a congruence transformation $B = X^TAX$. { Toomas page 145 }

Theorem: (**Cholesky Decomposition**). Suppose A is positive definite AND symmetric. Then you can write $L = GG^T$ where G is lower (or upper) triangular with positive diagonal elements.

Proof: Since A is pos def, we can write that $A = LDL^T$ from our theorem above, where L has unit diagonal. Since A is pos def, so is D, so D has positive diagonal elements. We can then write D as the product $\Lambda*\Lambda$ where Λ has the square root of the diagonal elements of D. Note that $\Lambda = \Lambda^T$ since Λ is diagonal. Then we have $A = LDL^T = L\Lambda\Lambda L^T = (L\Lambda)(L\Lambda)^T = GG^T$. The matrix G is thus lower triangular just like L, and it has the positive diagonal elements of Λ as its diagonal elements.

****

COMMENTS ON DIAGONALIZATION

1. You can transform an arbitrary square matrix A by similarity to upper Hessenberg form by two methods: The Givens' method with a finite number of planar rotations, or the Householder method. In either case, the similarity is unitary. If the matrix A is symmetric, then the Hessenberg form is tridiagonal.

2. In either of the above methods, there is a very well-defined prescription for applying the rotations or the elementary reflectors, so a computer program can easily get you to Hessenberg form with a finite number of operations, for a matrix of any reasonable size (say up to 100,000 x 100,000).

3. In contrast, the Schur decomposition tells us there exists a unitary similarity which takes general matrix A to full triangular form with eigenvalues on the diagonal. But we don't know what this similarity is, there is no straightforward way to compute it, and this is why we need that QR algorithm.

4. Similarly, you can get to Jordan Normal Form by similarity, but again, there is no simple numerical formula for finding this similarity matrix.

5. In these last two comments, we are talking about the numerical world, where we have to worry about errors and where the char poly is not clean, where you won't really know if two eigenvalues are the same or a little different, etc etc. You have to use numerical methods.

6. One application is in computing the zeros of orthogonal functions. The recursion relation when written correctly gives a tridiagonal matrix. You grind on this with the QR algorithm to get the eigenvalues, and those are then related to the zeros of the orthogonal functions. In this method, you might compute 100 zeros all at once to 30 places of accuracy. The claim is that this is more efficient that doing Newton-Raphson 100 times for the 100 zeros.